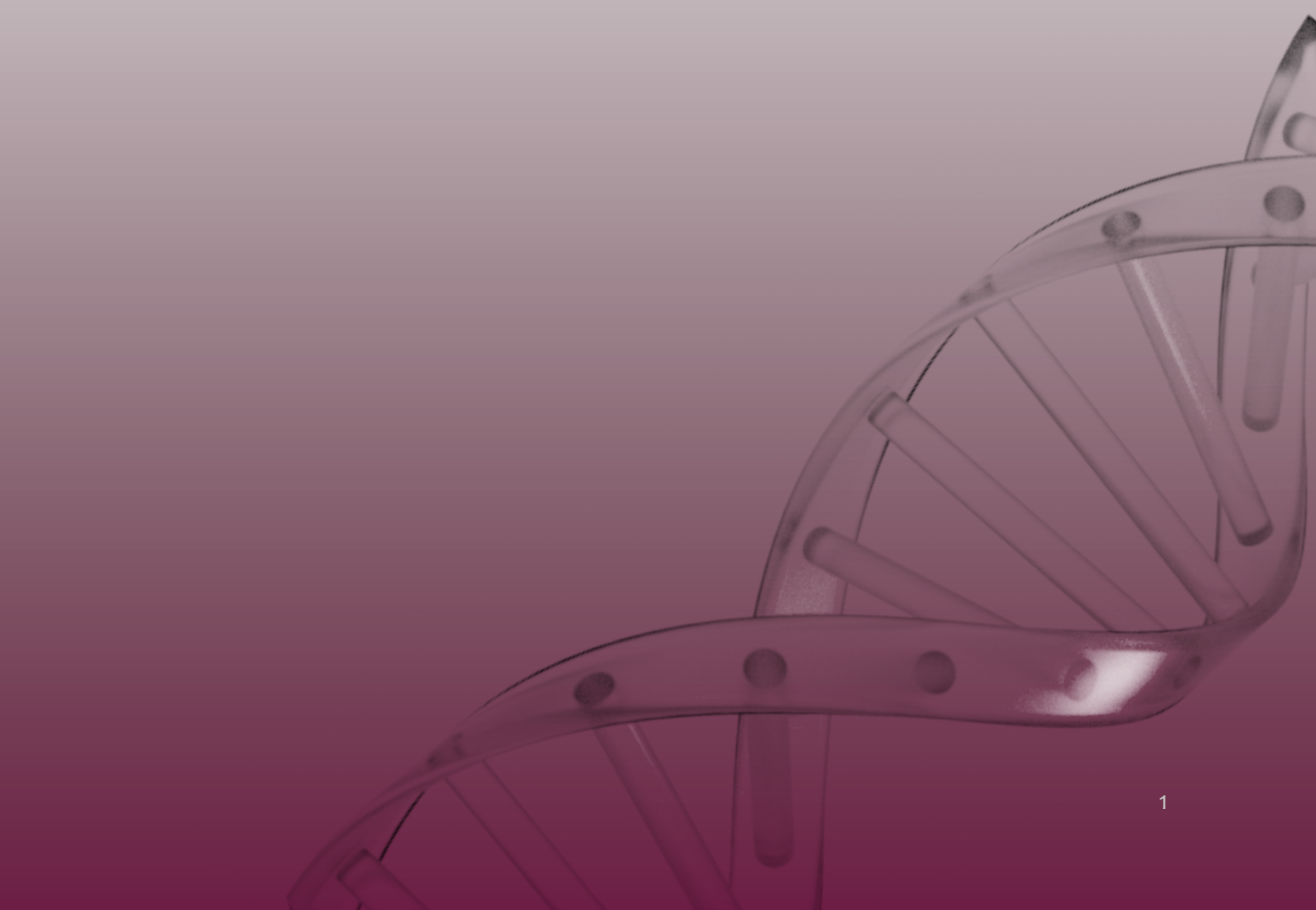
Less is more: Genome-reduced Bacillus subtilis for protein production

Aguilar Suarez, Rocio

DOI:
10.33612/diss.146898256

Chapter 1

Introduction



Reduced genomes in nature

Gene loss occurs naturally as a result of sudden mutation or as a slow process where mutations have acummulated[1,2]. This process can be observed in all kingdoms of life as part of evolution[3], and it occurs also in humans[4,5]. Evolution is not only the result of gene loss, but also gene duplication as observed in eukaryotes, such as plants and vertebrates[6]. During the evolution of vertebrates, humans acquired genes from bacteria through horizontal gene transfer. In fact, almost half of the human genome is derived from transposable elements, that fortunately have become inactive[7]. Intriguingly, some prokaryotes have experienced higher rates of gene loss than of gene acquisition, a trend that is observed particularly among genes belonging to the same family[8]. This loss of genes can compromise the free-living status of the respective organisms and facilitate intracellular lifestyles, ultimately leading to the development of parasites, endosymbionts and organelles. For instance, during evolution, chloroplasts transferred more than 3000 protein-encoding genes to their host. Thus, many proteins required for chloroplast function are synthetized by the host, and subsequently imported from the host's cytoplasm into the organelle. Chloroplasts have conserved ~100 protein-encoding genes which, overall, encode mostly proteins that are difficult to transport across the chloroplast's outer, inner and thylakoid membranes[9]. Chloroplasts and mitochondria are the current end points of an extreme reductive evolution that started with endo-symbiotic bacteria. For other microorganisms this evolution has reached an intermediate stage as exemplified by *Buchnera aphidicola.* This bacterium is an obligate symbiont, which lives in the gut of aphids. The presence of *B. aphidicola* is advantageous for the host, because it provides essential amino acids that are insufficiently provided through the diet of the aphids. Remarkably, *Buchnera* species evolved from the same ancestor as *Escherichia coli* (4.6 Mbp) and became an example of extreme genome reduction with genome sizes of approximately 600 kb[10]. Bacteria with intracellular lifestyles do not necessarily provide benefits to their hosts. *Wolbachia* species, which infect insects and transfer genes to their host, are able to modify the mating behaviour of the host by increasing female promiscuity to enhance their transmission to other hosts[11].

The examples above refer to organisms that have adapted to different conditions over several million years. However, genome reduction is not a process that exclusively occurs in nature since, in recent years, it has also been achieved in the laboratory by human efforts. This became possible as a result of the advances that have been achieved in genome sequencing and genome engineering in approaches that are collectively referred to as synthetic biology. Thus, the complete chemical synthesis of bacterial genomes has been achieved with the construction of *Mycoplasma mycoides* JCVI-syn1.0 and its minimized version JCVI-syn3.0 (531 kbp)[12,13]. Originally, genome reduction in the context of synthetic biology was primarily aimed at determining the minimal set of genes necessary to maintain cell viability and at elucidating gene functions[14,15]. However,

at the same time, it was realized that genome reduction could also offer the possibility to create novel chassis that could be exploited as bacterial cell factories[16,17].

Design-Build-Test cycle

The construction and improvement of genome-reduced microorganisms involves the Design-Build-Test cycle abbreviated as DBT cycle (Figure 1.1)[18]. The process begins with the design of a desired cell factory. Ideally, such a cell factory produces high yields of particular value-added products, utilizes low-cost substrates and presents an optimal productivity with minimal resource expenditure on biomass formation[19]. These preferred traits should be accompanied by low maintenance requirements in fermentation and high-level product secretion into the growth medium for ease of downstream processing of products. Some target sequences to be excluded from the final engineered genome include protease genes, transposons and insertion sequences, prophage regions and non-essential genes. Protease-deficient cell factories can improve the yields of protein products[20]. Deletion of non-essential genes, such as redundant genes, prophage regions or flagellar genes, may prevent the inefficient use of resources and enhance the stability of the producing cells. While, deletion of transposons and insertion sequences will increase their genetic stability[21]. The essentiality of genetic elements can be tested by inactivation through homologous recombination, transposon insertions, antisense RNA expression, or the CRISPRi technology[22,23]. Once the target genes for deletion have been defined, the following step in the process is to build the desired genome. However, in this context, it is necessary to consider the possibility of synthetically lethal genome configurations as a consequence of the simultaneous deletion of genes that are by themselves non-essential[24]. Lastly, the construction of the envisaged genome can be achieved along two major routes that involve either bottom-up or top-down approaches. The bottom-up approach requires chemical synthesis of DNA fragments and their consecutive assembly to create the desired genome[25]. Streamlining a genome through a top-down approach entails the successive excision of genomic segments[26]. To note, most of the genome-reduced microorganisms so far have been created following top-down approaches.

The next step in the DBT cycle requires to test the characteristics of the newly engineered strain. The performance of the strain should be evaluated to ensure that it meets the desired requirements. The acquired knowledge from this evaluation is fundamental for further improvement of the design. Since living cells display many different layers of regulation, whose intricate interactions are far from being completely understood, iterative DBT cycles will be required for fine-tuning and delivery of the ultimate cell factory.

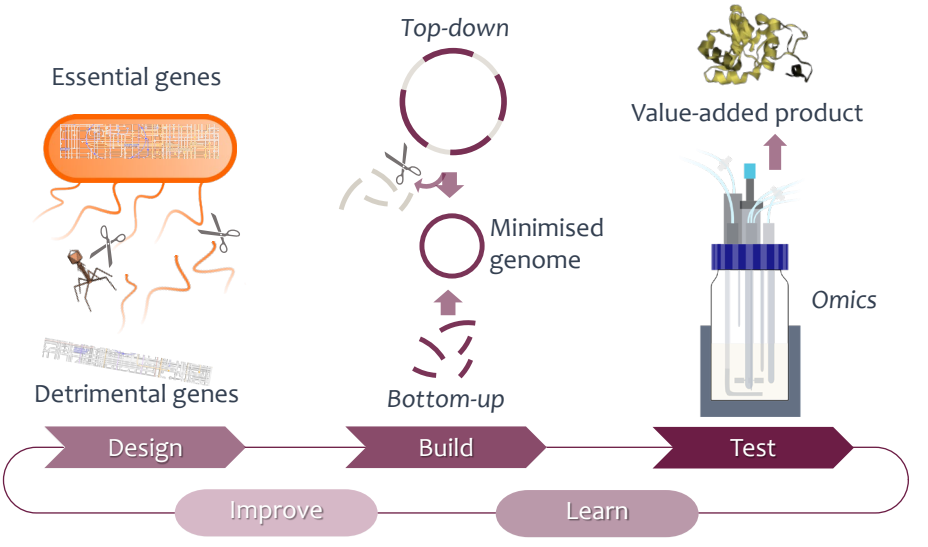


Figure 1.1 | Design–Build-Test cycle. Engineering of microorganisms by means of the DBT cycle. Properties that can be considered for the design of the cell factory include removal of non-essential or even detrimental genes that impose a metabolic burden on the cell. The second step involves the construction of the desired genome. To verify whether the design is functional, strains should be thoroughly tested. Finally, the acquired knowledge after the test stage is fundamental to learn and improve the design.

Microbial cell factories

Although the use of genome-reduced cell factories has been explored in different microorganisms, most efforts have addressed prokaryotic cells (Figure 1.2). The genome engineering of eukaryotic cells is currently also in progress, but it is complicated due to the cellular complexity as illustrated by a relatively modest genome reduction of ~5% in *Saccharomyces cerevisae*[27]. *E. coli* and *B. subtilis*, the main model organisms of Gram-negative and Gram-positive bacteria, respectively, are currently preferred targets for genome engineering. This has resulted in genome-reduced derivative *E. coli* strains that lack up to 36% of the genome, and *B. subtilis* strains that lack up to 42% of the genome[28,29]. Importantly, cell factories with reduced genomes have already shown beneficial traits compared to their parental strains. For instance, a derivative of *Streptomyces avermitilis*, lacking 18.5% of the genome, displays increased antibiotic production due to the selective deletion of genes involved in the production of the major endogenous secondary metabolites produced by the parental strain[30]. Also, the *E. coli* strain MGF-01 displayed increased threonine production after the removal of genes that are not required for growth in minimal media[31]. Unfortunately, the acquired properties of genome-reduced strains are not always beneficial for their application in biotechnology. Some genome-engineered strains display inferior growth compared to the respective parental strains, or show no improvement in productivity[32].

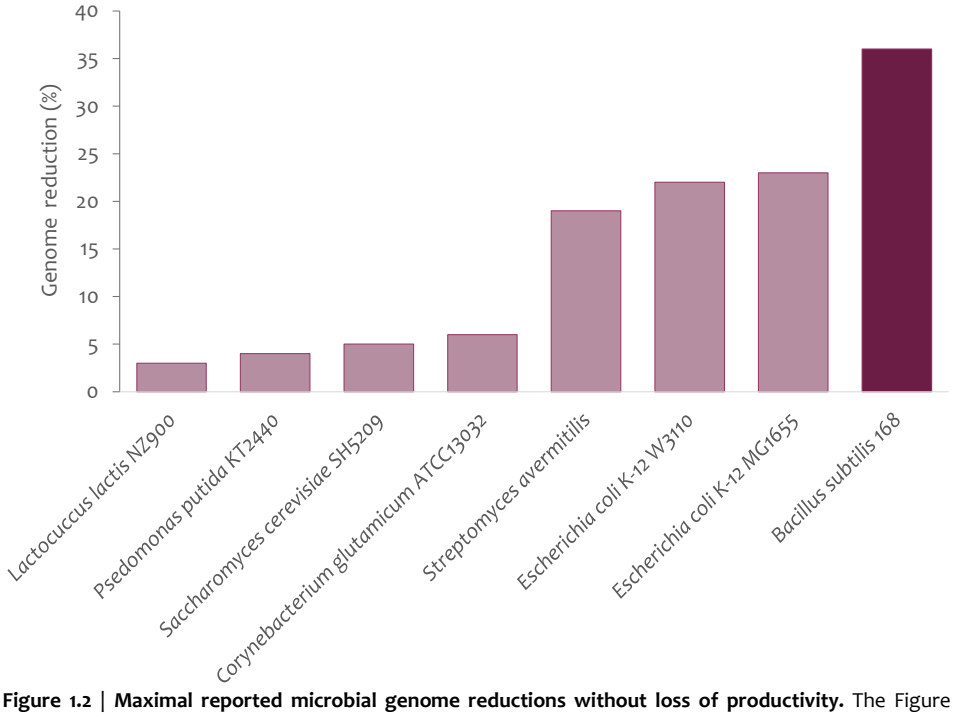


Figure 1.2 | Maximal reported microbial genome reductions without loss of productivity. The Figure indicates the names of the parental strains used in different top-down genome engineering efforts, where the resulting genome-reduced strains showed enhanced features for biotechnological applications. It should be noticed that in some of the indicated strains even larger sections of the genome have been deleted, but these have not yet been tested for application potential, or they showed no improvements in productivity. Therefore, the latter engineered strains were not considered for this comparison.

Bacillus subtilis as a cell factory

Bacillus species have been used extensively in industry for enzyme production, mainly because of their high secretion capacity and genetic amenability. Particularly, *B. subtilis* is a preferred cell factory as many of its products are 'Generally Recognised As Safe' by the Food and Drug safety Authority of the USA. The latter relates to the complete absence of toxins from *B. subtilis*, especially the lipopolysaccharides (endotoxins) produced by Gram-negative bacteria such as *E. coli*. However, a clear limitation of *B. subtilis* is its natural high-level production of proteases that can severely hamper the production of both homologous and heterologous proteins. Attempts to overcome this limitation resulted in the creation of the protease-deficient strains BRB08 and WB800, which lack the genes for the eight major extracellular proteases of *B. subtilis* 168[33,34]. From all industrially employed *Bacillus* species, genome reduction efforts have only been documented for *B. subtilis*. This is exemplified by the *B. subtilis* strain MG1M, which lacks 24% of the genome. However, the MG1M strain secreted a

cellulase and a subtilisin-like protease to similar levels as the parental 168 strain[35]. In contrast, another derivative of *B. subtilis* 168, namely strain MGB874 with a genome reduction of 20.7%, showed increased production of the secreted alkaline cellulase Egl-237 by about 2-fold[36].

A relatively recent systematic reduction of the *B. subtilis* 168 genome has resulted in the so far largest documented genome reduction of 42.3%[29]. The construction of this strain involved many consecutive deletion steps resulting in the delivery of a large collection of strains lacking different genomic regions. The PhD research described in this dissertation is built on investigations where different intermediate strains from this collection were characterised for their potential application in protein production. Major attention was focused on three genome-reduced strains in particular, which will be referred to in this thesis as midi*Bacillus*-I, midi*Bacillus*-II and mini*Bacillus*. The genomes of these strains have been reduced by approximately 26%, 31% and 35%, respectively, compared to the parental strain[26] (Figure 1.3).

Figure 1.3 | Phylogeny of the genome-reduced *B. subtilis* strains used in the present PhD research. *B. subtilis* strain 168 is the parental strain from which the Δ6, midi*Bacillus*-I, midi*Bacillus*-II and mini*Bacillus* strains were created. The respective genome reduction is indicated as a percentage of the genome of the parental strain 168.

The first steps towards this massive genome reduction in *B. subtilis* included a systematic gene function analysis and the subsequent rational deletion of five prophages and prophage-like regions, as well as the large *pks* operon[15,37]. Plasmid-based chromosomal integration-excision systems were used to delete the latter regions, resulting in the so-called Δ6 strain[37]. Prophage-free strains have led to promising results for industrial applications, as reported for *Corynebacterium glutamicum*[38]. The subsequent markerless *B. subtilis* strains with step-wise reduced genomes were constructed with the help of a counter-selection system based on the mannose phosphoenolpyruvate-dependent phosphotransferase system[39]. For this purpose,

manP, which encodes a mannose transporter, and *manA,* whose product is an isomerase, were deleted from the Δ6 strain. The *manP* gene, a spectinomycin resistance gene and the flanking sequences of the respective regions to be deleted were then inserted in a shuttle vector. To corroborate the integration into the chromosome, clones were first selected based on spectinomycin resistance. To check the subsequent excision of the plasmid, cells were grown on mannose, since these cells that are *manA*- and *manP*-deficient will be resistant to mannose. Additionally, a vector containing the *ccdB* gene under the control of a rhamnose promoter from *E. coli* was used for the construction of this series of strains with reduced genomes[26]. The genomic regions deleted in this process correspond to additional prophages regions, and genes required for the production of the antibiotics sublancin, subtilosin A, bacilysin and bacilysocin. Other deleted genes are involved in motility, metabolism of secondary carbon sources, sigma factors for sporulation, proteases genes, and genes of unknown function. As pointed out above, some bacterial features that are required for survival and competitive success in nature are unnecessary under production conditions, as exemplified by flagella. Other processes, like sporulation, are unwanted for industrial strains, as spores facilitate more readily the release of genetically modified strains into the environment, or allow isolation of valuable production strains from commercialized products. Moreover, considering that protein synthesis is an expensive process, the deletion of unnecessary or redundant genes could decrease the respective energy demands and would allow the re-routing of resources towards products[40].

Staphylococcal antigens as model proteins

As model proteins for production in the genome-reduced *B. subtilis* strains, the focus in the present PhD research was directed towards antigens from the human and livestock pathogen *Staphylococcus aureus*. This Gram-positive bacterium has acquired many different antimicrobial resistances and is estimated to account for 700 000 deaths per year, making it one of the top-10 threats to global health[41]. One of the strategies to combat antibiotic resistant pathogens is vaccination, and for this purpose adequate antigens are needed to trigger and boost the immune system[42]. However, currently there is no clinically approved vaccine available to prevent infections by this pathogen[43]. In principle, *S. aureus* displays and secretes a broad repertoire of proteins that could be used as antigens for vaccination. Four of these potential antigens were selected as reporter proteins for the here presented studies, namely: the chemotaxis inhibitory protein (CHIPS), the staphylococcal complement inhibitor (SCIN), the nuclease (Nuc) and the immunodominant staphylococcal antigen A (IsaA). While these four proteins served primarily to assess improvements in secretion by genome-reduced *B. subtilis* strains, they may also have future potential applications as targets in anti-staphylococcal immunotherapies or as diagnostic markers[44–46].

Scope of this dissertation

Dear reader of this thesis,

When looking for strategies to improve secretory protein production, one could go from the optimization of culture and fermentation conditions, via the selection of optimal promoters, ribosomal-binding sites and signal peptides, or the editing of metabolic pathways, to genome engineering. However, the central question that was directly addressed in the present PhD research was whether the massive genome-reductions already achieved in the midi*Bacillus*-I, midi*Bacillus*-II and mini*Bacillus* strains would allow improved protein production, and if so, which mechanisms would be hidden behind such improvements. Upon reading this dissertation, you will realize that it focuses mostly on the third step of the DBT cycle. This goes in hand with the three main objectives of this thesis, which were: 1. benchmarking genome-reduced *B. subtilis* strains as potential cell factories for the production of staphylococcal antigens; 2. identifying beneficial or non-favourable traits presented by *B. subtilis* strains with reduced genomes; and 3. defining the mechanisms underlying the observed effects on protein productivity.

One of the main characters is presented in Chapter 2 of this thesis: the *B. subtilis* strain PG10, also known as mini*Bacillus.* It is worthy to note that mini*Bacillus* is the most genome-reduced Gram-positive bacterium so far used as a cell factory. It lacks 35% of the genome compared to its parental strain *B. subtilis* 168. Notwithstanding this massive genome reduction, mini*Bacillus* outperformed its parental strain in the heterologous production and secretion of the four antigens IsaA, Nuc, CHIPS and SCIN from the human pathogen *S. aureus*.

Our second main character, in the order of appearance in this thesis but not in importance, is the *B. subtilis* strain IIG-Bs27-47-24, referred to as midi*Bacillus*-II. This strain belongs to the same phylogeny as mini*Bacillus*. However, its genome was only reduced by 31% compared to the parental strain 168. For the studies described in Chapter 3, a novel mass-spectrometry approach was exploited for absolute protein quantification, firstly to define the membrane proteome of midi*Bacillus*-II. Secondly, the approach was used to define changes in the membrane upon secretion of IsaA by midi*Bacillus*-II. The resulting quantitative dataset offered an unprecedented view of the secretion stress responses displayed by midi*Bacillus*-II.

To disclose major cellular adaptations and to point out possible differences displayed by midi*Bacillus*-II and its parental strain upon secretory protein production, the physiological changes in both strains were analysed through label-free protein quantification as described in Chapter 4. In addition to membrane proteins, also changes in the levels of cytosolic and extracellular protein were investigated for a comprehensive proteomic survey. The major changes are described in Chapter 4 with special focus on protein synthesis, stress responses and protein secretion.

Chapter 5 introduces the third main character in the present thesis, namely *B. subtilis* strain IIG27-39-1, the midi*Bacillus*-I. The midi*Bacillus*-I was also challenged to produce IsaA, but in a special setting. Here it is important to note that the characterization of *Bacillus* strains in a shake flask provides merely a 'sneak preview' of their potential for industrial application. Instead, cultivation in bioreactors offers the possibility to test the strain under controlled conditions that are more similar to the industrial settings[47]. midi*Bacillus*-I cells were therefore grown in laboratory-scale bioreactors, where several physiological parameters were measured and compared to the parental strain 168.

After reading about the characterization of the genome-reduced *B. subtilis* strains, especially mini*Bacillus*, midi*Bacillus*-I and midi*Bacillus*-II, you will arrive to the last part of this thesis, Chapter 6. Here, I present the main conclusions that can be drawn from the documented studies, and provide directions for future investigations towards genome-engineered *B. subtilis* production strains.

I hope you will enjoy reading this thesis,

September 2020

Rocío Aguilar Suárez